OPTIMAL DESIGNS FOR TWO-STAGE CLINICAL TRIALS WITH DICHOTOMOUS RESPONSES

by

Larry M. Pearson*

and

Donald A. Berry**

Technical Report No. 380

June, 1981

*Department of Mathematics, Statistics and Computer Science, Marquette University, Milwaukee, WI 53233.

**School of Statistics, University of Minnesota, Minneapolis, MN 55455.
This author's research supported in part by NSF Grant No. MCS80-01800.

OPTIMAL DESIGNS FOR TWO-STAGE CLINICAL TRIALS WITH DICHOTOMOUS RESPONSES

by

Larry M. Pearson*

and

Donald A. Berry**

ABSTRACT

Two-stage designs are considered for clinical trials involving two treatments with dichotomous responses. The first is the information-gathering stage; the treatment indicated to be the better from first-stage and prior data is used exclusively in the second stage. The objective is to maximize the expected number of successes in the entire trial. The length of the first stage is allowed to be arbitrary and fixed in advance, or to be optimized as a function of prior information and the patient horizon. In both cases the numbers of patients allocated to the two treatments in the first stage are to be optimized. Two forms of prior information are considered: the success probabilities are both known but which treatment has the larger of the two is not known, and one success probability is known while the other has an arbitrary distribution.

Key words and phrases: Clinical trials, two-stage designs, dichotomous responses, optimal treatment allocations, maximizing successes, two-armed bandits.

*Department of Mathematics, Statistics and Computer Science, Marquette University, Milwaukee, WI 53233.

**School of Statistics, University of Minnesota, Minneapolis, MN 55455.

This author's research supported in part by NSF Grant No. MCS80-01800.

1. INTRODUCTION

Patients in a clinical trial are assigned treatments from among those under consideration. The assignments are usually made randomly (cf. Simon 1977): each patient is assigned a treatment according to some predetermined probability distribution. But there is a large literature (Bather 1981, Berry 1972, 1978, Feldman 1962, Petkau 1978, Simon, Weiss, and Hoel 1975, Zelen 1969) in which assignments are made sequentially, depending on accumulating information about the various treatments.

According to Bailar (1976) and Simon (1977), sequential designs are not used in practice, and for a variety of reasons. While the overriding reason seems to be philosophical, there are mundane problems which result from the dependence of sequential procedures on "current information". Most sequential designs allocate according to the results of all previous allocations and so require instantaneous responses. There are few actual trials in which the possibility of making continual modifications of the design is present -- for example, no multicenter trials have this ability. More typically, trials proceed in stages, with a committee of clinicians and biometricians deciding on changes in protocol at periodic intervals. Even if the possibility of continual modification existed, the logistical problems involved might be insurmountable.

Much of the relevant literature (Anscombe 1963, Armitage 1975, Chernoff and Petkau 1981, Colton 1963, Cornfield, Halperin, and Greenhouse 1969, Day 1969) concerns two-stage trials, the first stage of which involves experimentation with two treatments, and normally distributed responses. More recently, Langenberg and Srinivasan (1981) consider two procedures

for splitting the first stage to allow for response delays. These papers all assume gain to be relative, depending only on the difference in the mean treatment response.

In this paper we consider two-stage trials in which responses are dichotomous. Responses from the first stage are assumed to be available in determining the allocation for the second stage. This assumption is only approximately correct in trials where stages follow each other immediately, for not all of the recent patients will have responded. In such trials, modifications of the procedures considered here will be evident, but optimality may be lost. While response time should affect the number and lengths of stages in an actual trial, these will be assumed independent for convenience. In particular, we assume that the total number of patients in the trial, N, is fixed. We consider two circumstances for the length n of the first stage: first, n is fixed, and second, n is optimized.

The objective of the trial is to maximize the expected proportion of successes in the entire trial; the worth of a procedure is the expected success proportion when following that procedure. A procedure with maximal worth is said to be optimal. This focus on number of successes in the trial rather than on the difference between the effectiveness in the two treatments seems more natural -- in addition, ideas generalize to more than two treatments.

Observations on the two treatments, labeled 1 and 2, can be represented by Bernoulli processes with success probabilities p_1 and p_2. The parameters p_1 and p_2 are not precisely known and information concerning the treatments can be formulated in terms of a joint prior distribution measure, F, on p_1 and p_2. We assume that observations

on each treatment are exchangeable so that at any stage the number of successes and failures on the two treatments in the previous stages, denoted by $(s_1,f_1;s_2,f_2)$, are sufficient statistics for (p_1,p_2). According to Bayes's theorem, the posterior measure of (p_1,p_2) is proportional to

$$p_1^{s_1}(1-p_1)^{f_1}p_2^{s_2}(1-p_2)^{f_2}dF(p_1,p_2).$$

A procedure for allocating treatments in the trial must specify, at each stage, the number of patients assigned to each treatment. Since the patients are assumed to be exchangeable any allocation of these numbers within the stage is acceptable, however, to lessen the possibility of bias the allocations should be randomized within each stage. Following an optimal procedure, the treatment with the larger probability of success, given by the maximum of the current means of the p_i, would be used exclusively in the second and final stage. Therefore, a procedure can be specified by n_1 and n_2, the numbers of observations on the two treatments in the first stage. Whether or not $n = n_1 + n_2$ is to be optimized, the corresponding worth is given by $W(n_1,n_2;N;F)$, or more simply,

$$(1.1) \qquad W(n_1,n_2) = (n_1Ep_1 + n_2Ep_2)/N + E\max\{E(p_1,p_2)|X_1,X_2\}(N-n)/N$$

where the initial distribution F is given throughout and X_i is the number of successes in the n_i observations on treatment i in the first stage. The joint distribution of X_1 and X_2 is given by

$$P[(X_1,X_2) = (i,j)|F] = \binom{n_1}{i}\binom{n_2}{j} E(p_1^i(1-p_1)^{n_1-i}p_2^j(1-p_2)^{n_2-j}|F)$$

for $i = 0,\ldots,n$, $j = 0,\ldots,n$.

Example 1.1. Assume a "flat" or uniform joint density in the unit square given by $dF(p_1,p_2) = dp_1dp_2$. Suppose $N = 10$ and $n = 5$ (which may not be an optimal first-stage length). For treatment allocation (n_1,n_2) we have

$$P[(X_1,X_2) = (i,j)|F] = (n_1+1)^{-1}(n_2+1)^{-1}$$

and

$$E\max\{E(p_1,p_2)|(X_1,X_2)\} = E\max\left\{\frac{X_1+1}{n_1+2}, \frac{X_2+1}{n_2+2}\right\}.$$

Hence the worth of (n_1,n_2) is

$$W(n_1,n_2) = \frac{1}{4} + \frac{1}{2(n_1+1)(n_2+1)} \sum_{i=0}^{n_1}\sum_{j=0}^{n_2} \max\left\{\frac{i+1}{n_1+2}, \frac{j+1}{n_2+2}\right\}$$

The relevant calculations yield: $W(5,0) = .55357$, $W(4,1) = .55833$, and $W(3,2) = .56250$. Since p_1 and p_2 are exchangeable in this example, allocation (3,2) and (2,3) are both optimal (when $n = 5$).□

The worth of any allocation is a convex combination of the probabilities of success in the first and second stages. It is evident from (1.1) that the second probability is not smaller than the first. Therefore, if the second stage is lengthened while the first stage is unchanged the worth is not decreased. This means that, for all n_1, n_2, N, F, and $M \geq 0$,

$$W(n_1,n_2;N;F) \leq W(n_1,n_2;N+M;F),$$

and in turn implies the following theorem.

THEOREM 1.1. Whether $n = n_1+n_2$ is fixed or optimized, for all F and $M \geq 0$,

$$\max_{n_1,n_2} W(n_1,n_2;N;F) \leq \max_{n_1,n_2} W(n_1,n_2;N+M;F).$$

In each of the next two sections we consider two kinds of trials: the first and second stages are of predetermined lengths and the length of the first stage is to be optimized -- both with N fixed. In Section 2, F has a two-point distribution of a certain type. In Section 3, p_2 is known (but arbitrary) and p_1 has an arbitrary distribution.

2. DEPENDENT TREATMENTS

Feldman (1962) considers a special kind of prior distribution, F, under which p_1 and p_2 are dependent, one which assigns all its mass to two points: (α,β) and (β,α). He assumes that the N patients are treated sequentially with the results of all previous patients available before treating the present patient. He shows that a myopic procedure, one which assigns the treatment with the larger current (expected) probability of success to the present patient, is optimal. Kelley (1974) finds necessary and sufficient conditions for optimality of myopic procedures when the distribution is concentrated on two points.

In the current setting, a myopic allocation assigns all patients in the first stage to the treatment with the larger initial probability of success. We will show, by example, that myopic allocations are not always optimal whether or not the length of the first stage is optimized. However, as the trial length approaches infinity, the asymptotic worth of a myopic allocation will, in both cases, approach the asymptotic worth of an optimal allocation.

Using a slightly ambiguous notation, we let F denote the initial probability that $(p_1,p_2) = (\alpha,\beta)$. Also, let $F(s_1,s_2)$ denote the probability of (α,β) given s_i successes in n_i observations on treatment i. From Bayes's theorem

$$(2.1) \qquad F^{-1}(i,j) = 1 + \left(\frac{1-F}{F}\right)\left(\frac{\beta}{\alpha}\right)^{i-j}\left(\frac{1-\beta}{1-\alpha}\right)^{n_1-i-(n_2-j)} .$$

From (1.1) and (2.1), the worth of an allocation (n_1,n_2) is given by

$$(2.2)\qquad N\,W(n_1,n_2) = n_1[F\alpha+(1-F)\beta] + n_2[(1-F)\alpha+F\beta]$$

$$+ (N-n_1-n_2)\sum_{i=0}^{n_1}\sum_{j=0}^{n_2} P[(X_1,X_2) = (i,j)\,|F]$$

$$\cdot \max\{F(i,j)\alpha + (1-F(i,j))\beta,\ (1-F(i,j))\alpha + F(i,j)\beta\}$$

$$= N\beta + (\alpha-\beta)[n_1F + n_2(1-F)]$$

$$+ (N-n_1-n_2)(\alpha-\beta)\sum_{i=0}^{n_1}\sum_{j=0}^{n_2}\binom{n_1}{i}\binom{n_2}{j}$$

$$\cdot \max\ F\alpha^i(1-\alpha)^{n_1-i}\beta^j(1-\beta)^{n_2-j},\ (1-F)\alpha^j(1-\alpha)^{n_2-j}\beta^i(1-\beta)^{n_1-i}\}.$$

2.1. Fixed Length First Stage

When n is fixed an allocation is found by maximizing (2.2) over (n_1,n_2) for which $n_1+n_2 = n$. Optimal allocations and their worths for various α, β, n, N, and $F = 1/2$ are shown in Tables 2.1 and 2.2. These were calculated by enumeration of (2.2).

[Tables 2.1 and 2.2 about here]

The following theorems rely on the special nature of F and clarify the allocation problem somewhat.

THEOREM 2.1. For n fixed and two-point distributions on $\{(\alpha,\beta),(\beta,\alpha)\}$ with $0 \le \beta \le \alpha \le 1$ and $0 \le F \le 1$,

$$(2.3)\qquad W(n_1,n_2;N;F) = W(n_1,n_2;N;1-F)$$

for all $0 \le n_i \le n$, $n_1+n_2 = n$.

Proof. Immediate from considerations of symmetry, or from (2.2).□

When $\alpha = 1-\beta$, (2.1) becomes

$$F^{-1}(i,j) = 1 + \left(\frac{1-F}{F}\right)\left(\frac{1-\alpha}{\alpha}\right)^{2(i-j)+n_1-n_2}, \tag{2.4}$$

so that a success on treatment 1 has the same effect on the distribution of (p_1,p_2) as a failure on treatment 2. Furthermore, if $\alpha = 1-\beta$ and $F = 1/2$ then every allocation for the first stage is optimal:

THEOREM 2.2. For n fixed and two-point distributions on $\{(\alpha,1-\alpha),(1-\alpha,\alpha)\}$ with $1/2 \le \alpha \le 1$,

$$W(n_1,n_2;N;1/2) = W(n,0;N;1/2),$$

for all $0 \le n_i \le n$ such that $n_1+n_2 = n$.

Proof. It follows from (2.2) that

$$W(n_1,n_2;N;F) = 1-\alpha + (2\alpha-1)[n_1F + n_2(1-F)]/N \tag{2.5}$$

$$+ \frac{2\alpha-1}{2} \sum_{i=0}^{n_1} \sum_{j=0}^{n_2} \binom{n_1}{i}\binom{n_2}{j}$$

$$\cdot \max\{F\alpha^{i+n_2-j}(1-\alpha)^{n_1-i+j},\ (1-F)\alpha^{j+n_1-i}(1-\alpha)^{n_2-j+i}\}$$

and it is sufficient to show that

$$\sum_{i=0}^{n_1} \sum_{j=0}^{n_2} \binom{n_1}{i}\binom{n_2}{j} \max\{F\alpha^{i+n_2-j}(1-\alpha)^{n_1-i+j},\ (1-F)\alpha^{j+n_1-i}(1-\alpha)^{i+n_2-j} \tag{2.6}$$

$$= \sum_{i=0}^{n} \binom{n}{i} \max\{F\alpha^{i}(1-\alpha)^{n-i},\ (1-F)\alpha^{n-i}(1-\alpha)^{i}\}$$

for all (n_1,n_2) such that $n_1+n_2 = n$. Actually, (2.6) is needed only for $F = 1/2$ but the more general result will be used in the proof of Theorem 2.3.

The left side of (2.6) equals

$$\sum_{m=0}^{n} \sum_{i=0}^{m} \binom{n_1}{i}\binom{n_2}{m-i} \max\{F\alpha^m(1-\alpha)^{n-m},\ (1-F)\alpha^{n-m}(1-\alpha)^m\}.$$

The result follows since

$$\sum_{i=0}^{m} \binom{n_1}{i}\binom{n_2}{m-i} = \binom{n}{m}.\square$$

The next theorem says that myopic procedures are optimal for some rather special two-point distributions. So in these cases it is optimal to allocate all n patients in the first stage to the same treatment.

THEOREM 2.3. For n fixed and two-point distributions on $\{(\alpha,\beta),(\beta,\alpha)\}$, myopic allocations are optimal in each of three cases: $\alpha = 1-\beta \geq 1/2$, $\alpha = 1$, and $\beta = 0$.

Proof. For $\alpha = 1-\beta \geq 1/2$, it follows from (2.5) and (2.6) that the expected number of successes from the second stage is independent of the treatment allocation in the first stage. Therefore, a myopic allocation is optimal.

When $\alpha = 1$,

$$\begin{aligned} W(n_1,n_2) &= \beta + (1-\beta)[n_1F + n_2(1-F)]/N + (1-\beta)\max\{F\beta^{n_2},\ (1-F)\beta^{n_1}\}/2 \\ &\quad + \frac{1-\beta}{2}\left[\sum_{j=0}^{n_2-1}\binom{n_2}{j}F\beta^j(1-\beta)^{n_2-j} + \sum_{i=0}^{n_1-1}\binom{n_1}{i}(1-F)\beta^i(1-\beta)^{n_1-i}\right] \\ &= \beta + (1-\beta)[n_1F + (n-n_1)(1-F) + n(1+\max\{(1-F)\beta^{n_1},\ F\beta^{n-n_1}\})]/N, \end{aligned}$$

which is maximized by $n_1 = 0$ or n according as $F \leq 1/2$ or $F \geq 1/2$.

The case $\beta = 0$ follows by a similar argument.□

Although the worth of a myopic allocation usually compares favorably with the worth of an optimal allocation for (α,β) and F not considered in the above theorem, as the next example suggests, a myopic allocation is not always optimal.

Example 2.1. Suppose $\alpha = .8$ and $\beta = .4$. Then

$$W(6,0) = W(0,6) = .672192,$$

which is somewhat less than

$$W(5,1) = W(1,5) = .672704.$$

The only optimal allocations are (5,1) and (1,5).□

While not necessarily optimal, myopic allocations are asymptotically optimal, as the subsequent development will show. First, we find an upper bound for the worth of an optimal allocation and then show that as the trial length becomes large both the worth of the optimal and myopic allocations approach this upper bound.

THEOREM 2.4. For n fixed and two-point distributions on $\{(\alpha,\beta),(\beta,\alpha)\}$ with $0 \leq \beta \leq \alpha \leq 1$ and $0 \leq F \leq 1$,

$$(2.7) \qquad W(n_1,n_2) \leq [n_1(F\alpha+(1-F)\beta) + n_2(F\beta+(1-F)\alpha) + n\alpha]/N$$

for all $0 \leq n_i \leq n$ with $n_1+n_2 = n$.

Proof. The expected number of successes from the first stage is $n_1(F\alpha+(1-F)\beta) + n_2(F\beta+(1-F)\alpha)$. The maximum expected number of successes from the second stage is $n\alpha$ which would result if p_1 and p_2 were

to become known after the first stage -- if $F(X_1,X_2) = 0$ or 1 with probability one.□

Remark. Inequality (2.7) is strict if $0 < F < 1$ and either $\beta > 0$ or $\alpha < 1$.

The next result is a special case of a well-known phenomenon: posterior distributions converge strongly. It is a straightforward consequence of the strong law of large numbers.

LEMMA 2.1. For the two-point distributions on $\{(\alpha,\beta),(\beta,\alpha)\}$ with $0 \leq \beta < \alpha \leq 1$ and $0 < F < 1$,

$$\lim_{n\to\infty} P[F(X_1,X_2) \to 0 \text{ or } 1 \mid F] = 1.$$

THEOREM 2.5. For two-point distributions on $\{(\alpha,\beta),(\beta,\alpha)\}$, assume $0 \leq \beta < \alpha \leq 1$ and $0 < F < 1$. If $n_i \to \infty$ as $n \to \infty$, $i = 1, 2$, and $n_1 = o(n)$ if $F < 1/2$ and $n_2 = o(n)$ if $F > 1/2$, then

$$\lim_{n\to\infty} W(n_1,n_2) = [\alpha + \max\{F\alpha+(1-F)\beta,\ (1-F)\alpha+F\beta\}]/2.$$

Proof. Under the conditions of the theorem, the maximal expected number of successes from the second stage will approach α according to Lemma 2.1. Choosing $n_1 = o(n)$ if $F < 1/2$ or $n_2 = o(n)$ if $F > 1/2$ assures that the expected proportion of successes from the first stage equals $\max\{F\alpha+(1-F)\beta,\ (1-F)\alpha+F\beta\}$ in the limit.□

While this shows that myopic allocations are asymptotically optimal, it makes it clear that many other allocations are also asymptotically optimal.

2.2. Optimal Length of First Stage

The restriction that n is fixed in advance is now removed. An optimal allocation can be found by maximizing (2.2) over (n_1,n_2) such that $0 \le n_1 + n_2 \le N$. Tables 2.3 and 2.4 give optimal allocations and worths for the same combinations of α, β, N and F as considered in Tables 2.1 and 2.2, respectively.

[Tables 2.3 and 2.4 about here]

The symmetry conditions described in Theorems 2.1 and 2.2 still hold. In particular, for $\alpha = 1-\beta \ge 1/2$ and $F = 1/2$ an optimal allocation of the form $(n,0)$ can always be found.

Example 2.2. Let $\alpha = 1-\beta \ge 1/2$ and $F = 1/2$. For n odd,

$$(2.8)\qquad NW(n,0) = N(1-\alpha) + (2\alpha-1)\left[\frac{n}{2} + \frac{N-n}{2}\sum_{i=0}^{(n-1)/2}\binom{n}{i}\alpha^{n-i}(1-\alpha)^{i}\right]$$

$$+ (2\alpha-1)\frac{N-n}{2}\sum_{i=(n+1)/2}^{n}\binom{n}{i}\alpha^{i}(1-\alpha)^{n-i}$$

$$= \frac{N}{2} + (2\alpha-1)\frac{N-n}{2}\left[I_{\alpha}\left(\frac{n+1}{2},\frac{n+1}{2}\right) - I_{1-\alpha}\left(\frac{n+1}{2},\frac{n+1}{2}\right)\right],$$

where $I_{\theta}(p,q)$ is the incomplete beta function,

$$I_{\theta}(p,q) = [\Gamma(p+q)/(\Gamma(p)\Gamma(q))]\int_{0}^{\theta} t^{p-1}(1-t)^{q-1}dt.$$

Since

$$\frac{\partial}{\partial x}\int_{a}^{b} t^{x}(1-t)^{x}dt = \int_{a}^{b}[\ln(t) + \ln(1-t)]t^{x}(1-t)^{x}dt$$

for $x \geq 1$ and $0 \leq a < b \leq 1$, the value of n (considered temporarily to be real rather than integral) which maximizes (2.8) satisfies the equation

$$N = 2(E[\ln(T_1) + \ln(1-T_1)] - E[\ln(T_2) + \ln(1-T_2)])^{-1} + n,$$

where the random variables T_1 and T_2 have densities

$$f_i(t_i) \propto t_i^{(n-1)/2}(1-t_i)^{(n-1)/2}$$

for $t_1 \in (1-\alpha,\alpha)$ and $t_2 \in (0,1)$, and 0 otherwise. It follows that the optimal odd first stage size is decreasing in α for $\alpha \geq 1/2$.

The above argument can be modified to show that the optimal even first stage size is also decreasing in α for $\alpha \geq 1/2$.□

Tables 2.3 and 2.4 support the notion that the optimal size for the first stage is again decreasing in α for $\alpha \geq 1/2$ and $\alpha \geq \beta$. Also, the maximal worth is increasing in α but not necessarily increasing in β. The worth of an optimal allocation for the various values of α, β, and F can be compared to α, an upper bound which is attained in the limit (as $N \to \infty$) by any sequence of allocations where the corresponding first stage sizes, n_N, satisfy the conditions $n_N \to \infty$ and $n_N = o(N)$ as $N \to \infty$. This follows by generalizing the proof of Theorem 2.5.

A myopic allocation can be shown optimal for the three cases of Theorem 2.3 by an evident modification of the proof of that theorem. The tables suggest that myopic allocations compare favorably with optimal allocations. However, the loss from using a myopic allocation is not as great when the first stage has a moderate length (cf. Section 2.1). This is because the optimal length of the first stage is typically quite small.

3. ONE TREATMENT KNOWN

Assume that one of the treatments, say treatment 2, has a known probability, λ, of success. This is applicable when treatment 2 is a standard treatment about which much is known.

When patients are treated individually and the results of all previously treated patients are known before treating the current patient, Bradt, Karlin, and Johnson (1956) show that optimal procedures are partially characterized by the stay-on-a-winner rule: if the current treatment results in a success then it is optimal to use that treatment on the next patient as well. Berry and Fristedt (1979) consider discounting future patients; the value of the m^{th} patient is b_m, where $b_{m+1} \leq b_m$. They give necessary and sufficient conditions on the sequence $b_1, b_2, b_3, \ldots$ for optimal procedures to be characterized by the stay-on-a-winner rule for all F. However, explicit solutions are not possible for the fully sequential problem for any interesting discount sequences (e.g., neither finite horizon nor geometric) unless F is very special. In general, a backward induction is necessary.

Again, when there are two stages, a procedure can be specified by (n_1, n_2), the numbers of observations on the two treatments in the first stage.

For the purposes of this section, F can be regarded as a pair (F_1, λ) where F_1 is the prior distribution of p_1 and λ is the known value of p_2. Since p_2 is known a priori, no amount of experimentation on treatment 2 changes F. However, observations on treatment 1 provide worthwhile information: the expected proportion of successes in the

second stage increases if additional observations are allocated to treatment 1 in the first stage. In fact, if n is fixed, the allocation $(n,0)$ will be seen to be optimal whenever $Ep_1 \geq \lambda$. And when n is optimized, we will show that it is always optimal to allocate all n patients to treatment 1.

3.1. Fixed Length First Stage

In this section, $n = n_1+n_2$ is fixed in advance. As more generally, an optimal allocation can be found by maximizing (1.1).

The following example shows the interplay between the current probability of success and the information to be gained for use in obtaining future successes.

Example 3.1. Suppose n is fixed ≥ 1, p_1 has a two-point distribution on $\{0,1\}$ with $F_1(\{1\}) = p^*$, and λ is arbitrary. Then

$$W(n_1,n_2) = \begin{cases} [n_1p^* + n_2\lambda + (N-n_1-n_2)(p^*+(1-p^*)\lambda]/N, & \text{if } n_1 \geq 1 \\ \lambda, & \text{if } n_1 = 0. \end{cases}$$

Comparing these expressions, all optimal allocations are seen to satisfy the following relations, where $\gamma = \lambda(1+n(1-\lambda))$:

For $p^* > \lambda$, $n_1 = n$,

For $p^* = \lambda$, $n_1 \geq 1$,

For $p^* \neq \lambda$ and $p^* < \gamma$, $n_1 = 0$,

For $p^* \neq \lambda$ and $p^* = \gamma$, $n_1 = 0$ or 1,

For $\gamma < p^* < \lambda$, $n_1 = 1$.

Because the distribution of p_1 in this example is so very special, complete information concerning treatment 1 can be obtained with but one observation. Therefore, if it is optimal to use it more than once in the first stage its probability of success must be at least that of treatment 2, that is, $p^* \geq \lambda$; and if $p^* > \lambda$ then treatment 1 must be used exclusively in the first stage since its probability of success is greater. Furthermore, the information gained from an observation on treatment 1 means that using it may be wise even though p^* is quite small, depending on n. For example, when $\lambda = 1/2$, if $1/(n+2) < p^* < 1/2$ then it is optimal to use treatment 1 once in the first stage.

The loss in worth from fixing n is apparent here -- in fact, it follows from Theorem 3.4 that an optimal length of the first stage is 1, uniquely optimal provided $p^* > \lambda/(\lambda + N(1-\lambda))$.□

The following theorem gives an analogue of the stay-on-a-winner rule, though it is much weaker than the fully sequential version. Its proof is straightforward and is omitted.

THEOREM 3.1. Assume n is fixed. If an optimal allocation assigns at least one patient to treatment 1 in the first stage and every observation on treatment 1 results in success then it is optimal to assign all patients in the second to treatment 1.

Obviously, Theorem 3.1 does not hold with treatment 2 in place of treatment 1. For, the use of treatment 2 in the second stage is independent of its performance in the first stage.

A phenomenon seen in Example 3.1 holds more generally:

THEOREM 3.2. For n fixed, if $E(p_1|F_1) \geq \lambda$ then $(n,0)$ is an optimal allocation, and $(n,0)$ is the uniquely optimal allocation provided

$E(p_1|F_1) > \lambda$.

Proof. Since $Ep_1 \geq \lambda$, for any allocation (n_1,n_2) the expected success proportion in the first stage,

$$[(n-n_1)\lambda + n_1Ep_1]/n,$$

is nondecreasing in n_1. Therefore, to prove that $(n,0)$ is an optimal allocation, it is sufficient to show that the expected proportion of successes from the second stage is nonincreasing in n_1. Letting Z denote the first observation on treatment 1,

$$(3.1) \qquad E[\max\{E(p_1|Z,1-Z;F_1),\lambda\}|F_1]$$

$$= Ep_1 \max\{E(p_1|1,0;F_1),\lambda\} + (1-Ep_1)\max\{E(p_1|0,1;F_1),\lambda\}$$

$$= \max\{Ep_1^2,\lambda Ep_1\} + \max\{Ep_1-Ep_1^2,\lambda(1-Ep_1)\}$$

$$\geq \max\{Ep_1,\lambda\},$$

with strict inequality provided $Ep_1 > \lambda$ and F_1 is not a one-point distribution. This gives the desired result for $n = 1$. And since (3.1) holds for all F_1, the expected proportion of successes in the second stage in nonincreasing for $n_1 \in \{0,1,\ldots,n\}$.□

The next theorem says that if the trial size is increased, the number of patients assigned to treatment 1 in the first stage should not decrease. Its proof is not instructive and is omitted.

THEOREM 3.3. For n fixed, let (n_1,n_2) be an optimal allocation. If the number of patients in the second stage is increased then there exists an optimal allocation, (n_1^*,n_2^*), for the larger trial with

$n_1^* \geq n_1$ (and, therefore, $n_2^* \leq n_2$).

Remark. Example 3.1 provides many instances in which $n_1^* = n_1$.

Tables 3.1 and 3.2 specify optimal allocations when F_1 is a beta distribution:

$$(3.2) \qquad dF_1(x) \propto x^{a-1}(1-x)^{b-1}dx.$$

For the grid of λ-values chosen for these tables, as λ is increased, the proportion of patients assigned to treatment 2 does not decrease. But as the following example shows, the issue is rather complicated.

[Tables 3.1 and 3.2 about here]

Example 3.2. Suppose p_1 has a uniform density on $(0,1)$, $0 \leq \varepsilon < 1/3$, and $n = 2$. If $\lambda = 1/2 - \varepsilon$ then (cf. Theorem 3.2)

$$N[W(2,0)-W(1,1)] = \left(\frac{N}{6} - \frac{4}{3}\right)\varepsilon ,$$

which is positive for $\varepsilon > 0$ and $N > 8$. So for $N > 8$, $n_1 = 1$ is optimal (but not uniquely) for $\lambda = 1/2$ but not optimal for $\lambda = 1/2 + \varepsilon$.□

A clear message from Tables 3.1 and 3.2 is that one treatment or the other tends to dominate the first stage (as well as the second -- but different treatments may dominate in different stages, depending on the results of the first). For example, the least extreme allocation on Table 3.2 is (12,38). This suggests that $n = 50$ is too large when $N = 100$, and this suggestion leads to the next topic.

3.2. Optimal Length First Stage

When the length of the first stage is given, Theorem 3.2 applies to show that an optimal allocation assigns all patients in the first stage to

treatment 1 provided $Ep_1 \geq \lambda$. This result can be easily extended to the case in which n is to be optimized. The next theorem has great intuitive appeal since the first stage is the information-gathering stage and there is nothing to learn about treatment 2 when p_2 is known.

THEOREM 3.4. For all $N \geq 1$, there exists an n, $0 \leq n < N$, such that $(n,0)$ is an optimal first stage allocation.

Proof. Suppose (n_1,n_2) is optimal. Then $(n_1,0)$ is also optimal since

$$(3.3) \quad NW(n_1,0;N;F) = n_1Ep_1 + (N-n_1)E[\max\{E(p_1|(X_1,n_1-X_1;F_1)),\lambda\}|F_1]$$

$$\geq n_1Ep_1 + n_2\lambda + (N-n_1-n_2)E(\max\{E(p_1|(X_1,n_1-X_1;F)),\lambda\}|F_1)$$

$$= NW(n_1,n_2;N;F).\square$$

Two particular situations where the treatment that is optimal for the second stage can be determined without computing the posterior probability of success for treatment 1, given the results of the first stage, are presented next. The first theorem is a variant of the stay-on-a-winner rule (cf. Theorem 3.1) and gives a switch-on-a-loser rule.

THEOREM 3.5. Let $N \geq 2$. Assume that a first-stage allocation of the form $(n,0)$ with $n > 0$ is uniquely optimal. Then, treatment 1 is optimal for the second stage if all successes are obtained in the first stage and treatment 2 is optimal if all failures are obtained.

Proof. If all successes are obtained in the first stage and treatment 1 is not optimal for the second stage then

$$(3.4) \quad E(p_1|(n,0;F_1)) < \lambda.$$

In view of (3.3) and (3.4),

$$W(n,0) = [n \cdot Ep_1 + (N-n)\lambda]/N$$

$$< \lambda = W(0,0),$$

which contradicts the optimality of the allocation $(n,0)$.

The other result follows by a similar argument.□

Example 3.3. Suppose p_1 has a uniform density on $(0,1)$: beta with $a = b = 1$ (cf. Table 3.3 and 3.4). From calculations similar to those in Example 1.1 and using $[\cdot]$ to denote the integer part, it follows that

$$N\,W(n,0) = \frac{n}{2} + \frac{N-n}{n+1} \sum_{i=0}^{n} \max\left\{\frac{i+1}{n+2}, \lambda\right\}$$

$$\doteq \frac{n}{2} + \frac{N-n}{n+1}\left(\sum_{i=0}^{[\lambda(n+2)-1]} + \sum_{i=[\lambda(n+2)}^{n} \frac{i+1}{n+2}\right)$$

$$\doteq \frac{n}{2} + \frac{N-n}{n+1}\left(\lambda^2(n+2) + \frac{n+1}{2} - \frac{\lambda(\lambda(n+2)+1}{2}\right)$$

$$= \frac{N}{2} + \frac{N-n}{2(n+1)}(\lambda^2(n+2) - \lambda)\,.$$

Differentiation gives

$$n^* = \{(N+1)(\lambda^{-1} - 1)\}^{1/2} - 2$$

as an approximation to the optimal first stage size.□

In the above example, n^* is nondecreasing in N. The next theorem says that the optimal first stage size in nondecreasing in N. It is a reasonably straightforward consequence of Theorem 3.4.

THEOREM 3.6: Suppose that $N < N'$. If $(n_1,0)$ is an optimal allocation for a trial of size N, then there exists an $n_1' \geq n_1$ such that $(n_1',0)$

is an optimal allocation for a trial of size N'.

Remark. Again, Example 3.1 provides many instances in which $n_1' = n_1$. Tables 3.3 and 3.4 also provide such instances, and cases in which $n_1' > n_1$ as well.

A generalization of the two-point prior distribution considered in Example 3.1 is provided in the following example.

Example 3.4. Suppose for $\varepsilon > 0$ that F_1 is a two-point distribution on $\{0,1-\varepsilon\}$ with $F_1(\{1-\varepsilon\}) = r$, $\lambda < 1-\varepsilon$, and $0 < r < 1$. If n observations are allocated to treatment 1 in the first stage, then, where X_1 continues to denote the number of successes in the first stage,

$$(X_1, n-X_1; F_1)(\{1-\varepsilon\}) = 1 \quad \text{for} \quad X_1 \geq 1,$$

$$(0, n; F_1)(\{1-\varepsilon\}) = 1 - [1 + (r/(1-r))\varepsilon^n]^{-1},$$

and

$$P(X_1 = 0 | F_1) = 1 - r + r\varepsilon^n.$$

Therefore, since the expected number of successes in the first stage, given at least one success, is $1 + (n-1)(1-\varepsilon)$,

$$\begin{aligned} N W(n,0) &= [1 + (n-1)(1-\varepsilon) + (N-n)(1-\varepsilon)]r(1-\varepsilon^n) \\ &\quad + (N-n)\max\{(1-\varepsilon)(1 - [1 + r\varepsilon^n/(1-r)]^{-1}), \lambda\}[1 - r(1-\varepsilon^n)] \\ &= r(1-\varepsilon)[n + (N-n)(1-\varepsilon^n)] + (N-n)[1 - r(1-\varepsilon^n)] \\ &\quad \cdot \max\{(1-\varepsilon)(1 - [1 + r\varepsilon^n/(1-r)]^{-1}), \lambda\}. \end{aligned}$$

However, if

$$(1-\varepsilon)(1 - [1 + r\varepsilon^n/(1-r)]^{-1}) > \lambda,$$

then the worth of the allocation $(n,0)$ is the same as the worth of the allocations $(N,0)$ and $(0,0)$. Thus, an optimal allocation and corresponding worth can be found by maximizing

$$r(1-\varepsilon)n + (N-n)[r(1-\varepsilon^n)(1-\varepsilon-\lambda) + \lambda].$$

In particular, for $\lambda = 1/2$, $N = 100$, $r \in \{.25, .5, .75, .95\}$, and $1-\varepsilon \in \{.6, .75, .9\}$ an optimal first stage size is 6.□

Tables 3.3 and 3.4 give optimal first stage sizes for $N = 50$ and 100 and various values of λ and various beta distributions with parameters (a,b) for p_1. If $Ep_1 = a/(a+b)$ is held constant, then the size of the first stage seems to decrease in $a + b$ for $Ep_1 < \lambda$ and increase for $Ep_1 > \lambda$.

[Tables 3.3 and 3.4 about here]

Throughout this paper the patient horizon N has been assumed given. That optimal allocations and worths may be robust with respect to N can be seen by comparing Tables 3.3 and 3.4 (and also Tables 2.3 and 2.4). Theorem 3.6 says that the number of patients allocated to treatment 1 in the first stage for $N = 100$ cannot be less than when $N = 50$ (whenever both are unique). But even when the optimal allocations are different the corresponding worths are not very different: the maximal difference in Tables 3.3 and 3.4 is 0.01022 -- for $(a,b) = (1/2,1/2)$ and $\lambda = 0.8$.

Table 2.1. Optimal values of n_1, n_2 and corresponding worths for $F = 1/2$, $n = 25$ and $N = 50$

β \ α	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
0	25 0 .07338	25 0 .14985	25 0 .22499	25 0 .30000	25 0 .37500	25 0 .45000	25 0 .52500	25 0 .60000	25 0 .67500	25 0 .75000
.1		25 0 .16324	25 0 .24058	25 0 .32029	23 2 .39833	23 2 .47459	23 2 .54994	25 0 .62500	25 0 .70000	25 0 .77500
.2			24 1 .26104	22 3 .33690	25 0 .41745	17 8 .49698	24 1 .57421	25 0 .64989	25 0 .72500	25 0 .80000
.3				19 6 .36012	22 3 .43515	22 3 .51610	24 1 .59651	24 1 .67421	23 2 .74994	25 0 .82500
.4					22 3 .45973	25 0 .53462	22 3 .61610	17 8 .69698	23 2 .77459	25 0 .85000
.5						22 3 .55973	22 3 .63515	25 0 .71745	23 2 .79833	25 0 .87500
.6							19 6 .66012	22 3 .73690	25 0 .82029	25 0 .90000
.7								24 1 .76104	25 0 .84058	25 0 .92499
.8									25 0 .86324	25 0 .94985
.9										25 0 .97338

Table 2.2. Optimal values of n_1, n_2 and corresponding worths for $F = 1/2$, $n = 50$ and $N = 100$

	α									
β	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
0	50 0 .07488	50 0 .15000	50 0 .22500	50 0 .30000	50 0 .37500	50 0 .45000	50 0 .52500	50 0 .60000	50 0 .67500	50 0 .75000
.1		44 6 .16772	48 2 .24678	48 2 .32434	50 0 .39992	50 0 .47500	48 2 .55000	47 3 .62500	50 0 .70000	50 0 .77500
.2			50 0 .26478	48 2 .34429	45 5 .42343	49 1 .49977	46 4 .57498	50 0 .65000	47 3 .72500	50 0 .80000
.3				50 0 .36363	47 3 .44289	39 11 .52289	49 1 .59967	46 4 .67498	48 2 .75000	50 0 .82500
.4					50 0 .46315	49 1 .54224	39 11 .62289	49 1 .69977	50 0 .77500	50 0 .85000
.5						50 0 .56315	47 3 .64289	45 5 .72343	50 0 .79992	50 0 .87500
.6							50 0 .66363	48 2 .74429	48 2 .82434	50 0 .90000
.7								50 0 .76478	48 2 .84678	50 0 .92500
.8									44 6 .86722	50 0 .95000
.9										50 0 .97488

Table 2.3. Optimal values of n_1, n_2 and corresponding worths for $F = 1/2$ and $N = 50$

β \ α	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
0	14 0 .07859	8 0 .16926	7 0 .27156	5 0 .37160	4 0 .47281	3 0 .57478	2 0 .67693	2 0 .78093	1 0 .88659	1 0 .99000
.1		13 2 .16492	13 0 .24911	11 0 .34261	9 0 .44058	8 0 .54154	5 1 .64435	3 2 .74909	3 0 .85494	1 0 .98659
.2			15 1 .26237	13 1 .34344	11 1 .43535	7 3 .53373	8 0 .63601	7 0 .74079	3 2 .84909	2 0 .98093
.3				13 3 .36125	14 0 .44093	11 1 .53270	11 0 .63159	8 0 .73601	5 1 .84435	2 0 .97693
.4					13 3 .46080	14 1 .54017	11 1 .63270	7 3 .73373	8 0 .84154	3 0 .97478
.5						13 3 .56080	14 0 .64093	11 1 .73535	9 0 .84058	4 0 .97281
.6							13 3 .66125	13 1 .74344	11 0 .84261	5 0 .97160
.7								15 1 .76237	13 0 .84911	7 0 .97156
.8									13 2 .86492	9 0 .97320
.9										14 0 .97859

Table 2.4. Optimal values of n_1, n_2 and corresponding worths for F = 1/2 and N = 100

β \ α	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
0	20 0 .08562	13 0 .18317	9 0 .28264	7 0 .38288	5 0 .48379	4 0 .58505	3 0 .68675	2 0 .78886	1 0 .89105	1 0 .99500
.1		26 2 .16996	20 1 .26113	16 0 .35917	13 0 .46003	11 0 .56186	8 1 .66456	6 0 .76821	3 2 .87349	1 0 .99105
.2			30 0 .26678	23 1 .35539	19 0 .45314	12 3 .55454	10 2 .65774	8 1 .76231	6 0 .86821	2 0 .98886
.3				30 0 .36530	22 3 .45273	15 5 .55057	14 1 .65300	10 2 .75774	8 1 .86456	3 0 .98675
.4					30 0 .46478	24 1 .55193	15 5 .65057	12 3 .75454	11 0 .86186	4 0 .98505
.5						30 0 .56478	22 3 .65273	19 0 .75314	13 0 .86003	5 0 .98379
.6							30 0 .66536	23 1 .75531	16 0 .85917	7 0 .98288
.7								30 0 .76678	20 1 .86113	9 0 .98264
.8									26 2 .86996	13 0 .98317
.9										20 0 .98562

Table 3.1. Optimal values of n_1, n_2 and corresponding worths for $n = 25$ and $N = 50$ when $p_2 = \lambda$ is known and p_1 has a beta distribution (3.2)

	λ									
(a,b)	.05	.1	.2	.3	.4	.5	.6	.7	.8	.9
(4,1)	25 0 .80000	25 0 .80000	25 0 .80000	25 0 .80009	25 0 .80056	25 0 .80211	25 0 .80601	25 0 .81428	25 0 .82983	0 25 .90000
(1,½)	25 0 .66679	25 0 .66752	25 0 .67102	25 0 .67735	25 0 .68680	25 0 .69951	25 0 .71584	8 17 .74443	3 22 .81232	0 25 .90000
(2,1)	25 0 .66667	25 0 .66671	25 0 .66748	25 0 .67009	25 0 .67562	25 0 .68519	25 0 .70004	9 16 .72824	0 25 .80000	0 25 .90000
(½,½)	25 0 .50173	25 0 .50583	25 0 .51817	25 0 .53464	25 0 .55465	25 0 .57806	4 21 .64083	2 23 .71700	0 25 .80000	0 25 .90000
(1,1)	25 0 .50025	25 0 .50171	25 0 .50855	25 0 .52051	25 0 .53775	25 0 .56019	4 21 .62200	3 22 .70050	0 25 .80000	0 25 .90000
(2,2)	25 0 .50000	25 0 .50012	25 0 .50233	25 0 .50920	25 0 .52258	25 0 .54358	2 23 .60600	0 25 .70000	0 25 .80000	0 25 .90000
(4,4)	25 0 .50000	25 0 .50000	25 0 .50019	25 0 .50241	25 0 .51088	25 0 .52982	0 25 .60000	0 25 .70000	0 25 .80000	0 25 .90000
(1,2)	25 0 .33386	25 0 .33673	25 0 .34972	25 0 .37118	6 19 .41978	2 23 .50167	0 25 .60000	0 25 .70000	0 25 .80000	0 25 .90000
(½,1)	25 0 .33605	25 0 .34250	25 0 .36144	25 0 .38631	4 21 .43706	2 23 .51476	2 23 .60076	0 25 .70000	0 25 .80000	0 25 .90000
(1,4)	25 0 .20115	25 0 .20665	25 0 .22983	1 24 .30133	0 25 .40000	0 25 .50000	0 25 .60000	0 25 .70000	0 25 .80000	0 25 .90000

Table 3.2. Optimal values of n_1, n_2 and corresponding worths for $n = 50$ and $N = 100$ when $p_2 = \lambda$ is known and p_1 has a beta distribution (3.2)

	λ									
(a,b)	.05	.1	.2	.3	.4	.5	.6	.7	.8	.9
(4,1)	50 0 .80000	50 0 .80000	50 0 .80001	50 0 .80015	50 0 .80076	50 0 .80257	50 0 .80682	50 0 .81549	50 0 .83122	0 50 .90000
(1,½)	50 0 .66687	50 0 .66771	50 0 .67142	50 0 .67795	50 0 .68750	50 0 .70033	50 0 .71673	12 38 .74727	4 46 .81683	0 50 .90000
(2,1)	50 0 .66667	50 0 .66676	50 0 .66771	50 0 .67058	50 0 .67643	50 0 .68632	50 0 .70129	12 38 .73153	4 46 .80419	0 50 .90000
(½,½)	50 0 .50203	50 0 .50632	50 0 .51877	50 0 .53532	50 0 .55540	50 0 .57879	4 46 .64483	3 47 .72134	3 47 .80272	0 50 .90000
(1,1)	50 0 .50041	50 0 .50207	50 0 .50924	50 0 .52149	50 0 .53884	50 0 .56127	7 43 .62633	3 47 .70650	0 50 .80000	0 50 .90000
(2,2)	50 0 .50001	50 0 .50028	50 0 .50290	50 0 .51025	50 0 .52399	50 0 .54509	5 45 .61048	0 50 .70000	0 50 .80000	0 50 .90000
(4,4)	50 0 .50000	50 0 .50000	50 0 .50039	50 0 .50327	50 0 .51249	50 0 .53171	3 47 .60003	0 50 .70000	0 50 .80000	0 50 .90000
(1,2)	50 0 .33416	50 0 .33741	50 0 .35079	50 0 .37242	8 42 .42396	4 46 .50524	0 50 .60000	0 50 .70000	0 50 .80000	0 50 .90000
(½,1)	50 0 .33652	50 0 .34319	50 0 .36228	50 0 .38716	6 44 .44081	4 46 .51967	2 48 .60610	0 50 .70000	0 50 .80000	0 50 .90000
(1,4)	50 0 .20165	50 0 .20786	50 0 .23122	4 46 .30433	0 50 .40000	0 50 .50000	0 50 .60000	0 50 .70000	0 50 .80000	0 50 .90000

Table 3.3. Optimal values of n_1, n_2 and corresponding worths for $N = 50$ when $p_2 = \lambda$ is known and p_1 has a beta distribution (3.2)

	λ									
(a,b)	.05	.1	.2	.3	.4	.5	.6	.7	.8	.9
(4,1)	12 0 .80000	40 0 .80000	32 0 .80001	36 0 .80009	21 0 .80059	16 0 .80239	14 0 .80736	10 0 .81920	8 0 .84386	0 0 .90000
(1,½)	27 0 .66679	22 0 .66757	11 0 .67216	10 0 .68150	7 0 .69619	7 0 .71704	6 0 .74553	5 0 .78234	3 0 .83020	0 0 .90000
(2,1)	42 0 .66667	31 0 .66671	19 0 .66753	15 0 .67067	13 0 .67812	10 0 .69231	8 0 .71588	5 0 .75238	4 0 .80686	0 0 .90000
(½,½)	18 0 .50200	10 0 .50769	9 0 .52684	6 0 .55466	4 0 .58984	3 0 .63219	4 0 .68184	2 0 .74000	3 0 .80403	0 0 .90000
(1,1)	27 0 .50025	21 0 .50183	11 0 .51100	10 0 .52909	7 0 .55733	5 0 .59643	4 0 .64720	3 0 .71150	0 0 .80000	0 0 .90000
(2,2)	41 0 .50000	30 0 .50015	18 0 .50255	14 0 .51129	10 0 .53077	7 0 .56515	5 0 .61786	0 0 .70000	0 0 .80000	0 0 .90000
(4,4)	10 0 .50000	37 0 .50000	29 0 .50020	20 0 .50256	13 0 .51294	9 0 .54132	0 0 .60000	0 0 .70000	0 0 .80000	0 0 .90000
(1,2)	26 0 .33387	20 0 .33709	10 0 .35534	8 0 .38967	6 0 .44089	2 0 .50933	0 0 .60000	0 0 .70000	0 0 .80000	0 0 .90000
(½,1)	18 0 .33652	10 0 .34555	8 0 .37571	5 0 .41901	4 0 .47267	3 0 .53625	2 0 .61128	0 0 .70000	0 0 .80000	0 0 .90000
(1,4)	24 0 .20115	18 0 .20788	8 0 .24386	3 0 .30743	0 0 .40000	0 0 .50000	0 0 .60000	0 0 .70000	0 0 .80000	0 0 .90000

Table 3.4. Optimal values of n_1, n_2 and corresponding worths for $N = 100$ when $p_2 = \lambda$ is known and p_1 has a beta distribution (3,2)

	λ									
(a,b)	.05	.1	.2	.3	.4	.5	.6	.7	.8	.9
(4,1)	83 0 .80000	69 0 .80000	47 0 .80001	40 0 .80016	31 0 .80088	26 0 .80321	21 0 .80918	17 0 .82237	12 0 .84881	0 0 .90000
(1,½)	45 0 .66688	24 0 .66795	21 0 .67334	14 0 .68371	12 0 .69970	10 0 .72183	8 0 .75128	8 0 .78889	4 0 .83722	0 0 .90000
(2,1)	64 0 .66667	42 0 .66676	29 0 .66791	22 0 .67181	18 0 .68048	14 0 .69618	11 0 .72144	9 0 .75932	5 0 .81369	0 0 .90000
(½,½)	21 0 .50264	18 0 .50915	13 0 .52984	9 0 .55917	7 0 .59555	5 0 .63916	4 0 .68975	5 0 .74717	4 0 .81425	0 0 .90000
(1,1)	44 0 .50043	24 0 .50257	16 0 .51318	13 0 .53314	9 0 .56287	9 0 .60341	7 0 .65500	3 0 .71825	0 0 .80000	0 0 .90000
(2,2)	63 0 .50002	41 0 .50029	28 0 .50354	21 0 .51354	15 0 .53561	11 0 .57188	7 0 .62541	0 0 .70000	0 0 .80000	0 0 .90000
(4,4)	81 0 .50000	66 0 .50000	44 0 .50040	30 0 .50378	23 0 .51644	15 0 .54720	8 0 .60455	0 0 .70000	0 0 .80000	0 0 .90000
(1,2)	43 0 .33421	23 0 .33849	15 0 .35903	12 0 .39587	8 0 .44857	4 0 .51619	0 0 .60000	0 0 .70000	0 0 .80000	0 0 .90000
(½,1)	21 0 .33750	18 0 .34775	9 0 .38006	8 0 .42489	6 0 .47920	5 0 .54423	2 0 .61707	0 0 .70000	0 0 .80000	0 0 .90000
(1,4)	41 0 .20180	21 0 .21033	12 0 .24881	7 0 .31273	0 0 .40000	0 0 .50000	0 0 .60000	0 0 .70000	0 0 .80000	0 0 .90000

REFERENCES

Anscombe, F. J. (1963). Sequential medical trials. J. Am. Statist. Assoc. 58, 365-83.

Armitage, P. (1975). Sequential Medical Trials, 2nd edition, Oxford: Blackwell.

Bailar, J. C. (1976). Patient assignment algorithms -- an overview. In Proceedings of the 9th International Biometrics Conference, Volume I, Raleigh, North Carolina: The Biometric Society, 189-206.

Bather, J. A. (1981). Randomized allocation of treatments in sequential experiments. J. R. Statist. Soc. B 43, 000-000.

Berry, D. A. (1972). A Bernoulli two-armed bandit. Ann. Math. Statist. 43, 871-897.

Berry, D. A. (1978). Modified two-armed bandit strategies for certain clinical trials. J. Am. Statist. Assoc. 73, 339-345.

Berry, D. A. and Fristedt, B. (1979). Bernoulli one-armed bandits -- arbitrary discount sequences. Ann. Statist. 7, 1086-1105.

Chernoff, H. and Petkau, A. J. (1981). Sequential medical trials involving paired data. Biometrika 68, 119-132.

Colton, T. (1963). A model for selecting one of two medical treatments. J. Am. Statist. Assoc. 58, 388-400.

Cornfield, J., Halperin, M., and Greenhouse, S. W. (1969). An adaptive procedure for sequential clinical trials. J. Am. Statist. Assoc. 64, 759-770.

Day, N. E. (1969). A comparison of some sequential designs. Biometrika 56, 301-311.

Feldman, D. (1962). Contributions to the "two-armed bandit" problem. Ann. Math. Statist. 33, 847-856.

Langenberg, P. and Srinivasan, R. (1981). On the Colton model for clinical trials with delayed observations -- normally distributed responses. Biometrics 37, 143-148.

Petkau, A. J. (1978). Sequential medical trials for comparing an experimental with a standard treatment. J. Am. Statist. Assoc. 73, 328-345.

Simon, R. (1977). Adaptive treatment assignment methods and clinical trials. Biometrics 33, 743-749.

Simon, R., Weiss, G. H., and Hoel, D. G. (1975). Sequential analysis of binomial clinical trials. Biometrika 62, 195-200.

Zelen, M. (1969). Play-the-winner rule and controlled clinical trials. J. Am. Statist. Assoc. 64, 131-140.